GINAS: Global Ingredient Archival Service

TOWARD AN OPEN IMPLEMENTATION OF SUBSTANCE REGISTRATION AND ISO 11238

September 12, 2013

NCATS

NIH National Center for Advancing Translational Sciences

GInAS Need

- Global marketplace for ingredients requires a global system to monitor the global supply chain
- An ISO standard ISO 11238 has recently been developed to describe all substances in medicinal products
- Expensive to implement a system based on 11238 on a individual basis may prevent adoption of standard
- Pharmacovigilence based on substances with global data
- Better coordination of regulatory activity and clinical trials (inspections, specifications, drug shortages)
- Without a global system extensive mapping is required
- Standards can converge more rapidly

ISO 11238 and GInAS Scope

- Substance categories
  - Chemical
  - Protein
  - Nucleic acid
  - Polymer
  - Structurally diverse
- Specified substances Groups 1–4
- Official names in multiple languages, jurisdictions, and domains
- Well-defined references and relationships between substances (impurities, metabolites, targets etc.)
- Unique identifiers
- Documents and Data associated with substances

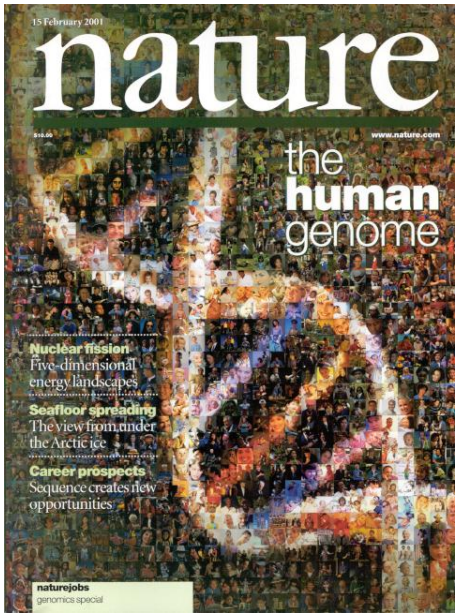
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=55031

GInAS

- Meeting Hosted by USP and NIH/NCATS on ISO 11238 substance registration standard
- Canadian, Dutch, German, Swiss, and US regulators, EDQM and USP involved in the development of the system
- Seed funding from NIH/NCATS and FDA (SRS development)
- Initial deployment hosted at Health Canada

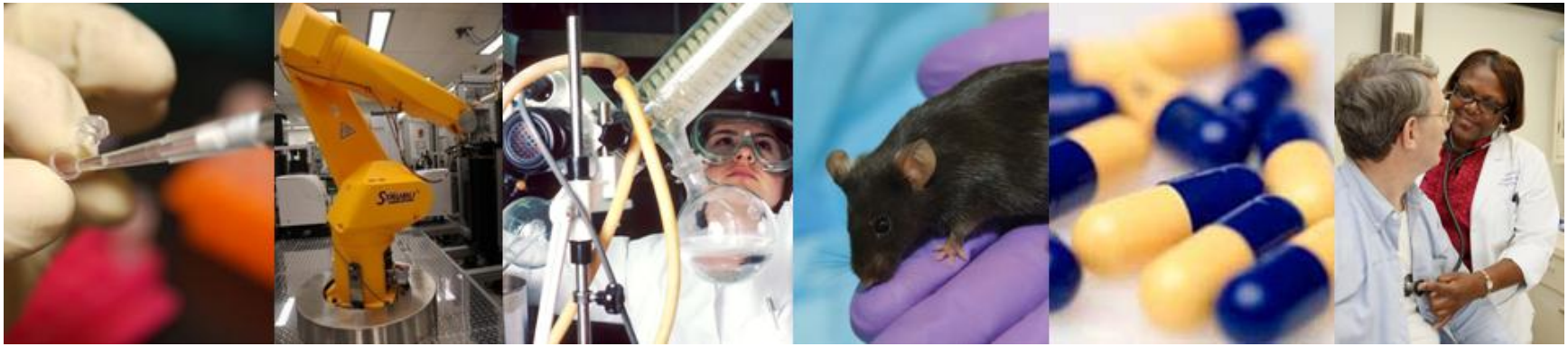
“Houston, we have a problem”

- Fundamental science unprecedentedly advanced, but:
  - Poor transition of those advances to interventions that tangibly improve human health
  - Drug/device development system in crisis
  - Clinical trials system in crisis
  - Poor adoption of demonstrably useful interventions



People unhealthier and funders of biomedical research enterprise (public and private) impatient

NCATS Mission



To catalyze the generation of innovative methods and technologies that will enhance the development, testing and implementation of diagnostics and therapeutics across a wide range of human diseases and conditions.

NCATS Mission

To catalyze the generation of innovative methods and technologies that will enhance the development, testing and implementation of diagnostics and therapeutics across a wide range of human diseases and conditions.

- Clinical Translational Science Awards
- Discovering New Therapeutic Uses for Existing Molecules
- Therapeutics for Rare and Neglected Disease
- Toxicology in the 21st Century (Tox21)
- NIH Chemical Genomics Center

Characteristics of NCATS Initiatives and Programs

- Address significant bottlenecks in the process of translation
- Highly collaborative across NIH, other government agencies, and with the private sector.
- Quick to respond to needs of biomedical researchers
- *Develop*, *demonstrate*, and *disseminate* software for managing, mining, and visualizing data

NCATS Informatics

- Unified pool of talent to support diverse range of missions, from next-generation sequencing to clinical data support
- Experience supporting mission critical applications at large organizations
- Development cycle iterative, in close collaboration with domain experts

GInAS/11238 Collaborators

BfArM (Germany) Thomas Balzer
CBG-MEB (Netherlands) Herman Diederik, Ciska Matai
EDQM Claude Coune (retired), Chris Jarvis
EMA Sabine Brosch, Tim Buxton, Ilaria Del Seppia, Paul Houston, Telonis Pangiotis
FDA (US) Yulia Borodina, Larry Callahan, Vada Perkins, Frank Switzer
Health Canada Vik Srivastava
NCATS (US) Trung Nguyen, Tyler Peryea, Noel Southall
PAHO/WHO Analia Porras
Swiss Medic Philipp Weyermann
USP Fouad Atouf, Tina Morris, Andrzej Wilk

GInAS Goals

- Develop and deploy an information system that can serve as a global repository for definitional, regulatory and scientific information on substances
- Establish a consortium of regulators and other international organizations to manage and govern the repository
- Develop and distribute a global identifier for every substance in marketed medicinal products and clinical research
- Distribute an information system to both regulators, companies and other interested parties to facilitate registration into the global repository

Development Strategy

- Core software developers with intimate domain knowledge and experience with regulatory process
- Iterative development cycle
  - » Functional requirements through in-depth case studies with users/experts

    http://tripod.nih.gov/pub/iso11238/case_studies/
  - » Discussion and feedback from case studies drive design and development
  - » Evaluate implemented features with users through usability studies
  - » ISO 11238 core standard; detailed system description/requirements necessitate a working implementation
- Bi-weekly iteration cycle

Implementation Goals for Software

- Self-contained and modular
  - » Run entirely on a desktop or access remotely
  - » Freely distributable, predominantly Open Source
- Well-defined data access application programming interface (API)
  - » GInAS or third-party clients / internal business processes
- Fine-gained security model
  - » Access control for every piece of information
  - » Audit trail of all data fields
- Configurable “business rules”, e.g. standardizing structures
- All data is referenced; primary sources should be retained (e.g., PDF’s, MS spectra, images)
- System will be distributed with a large set of public domain data and updated periodically

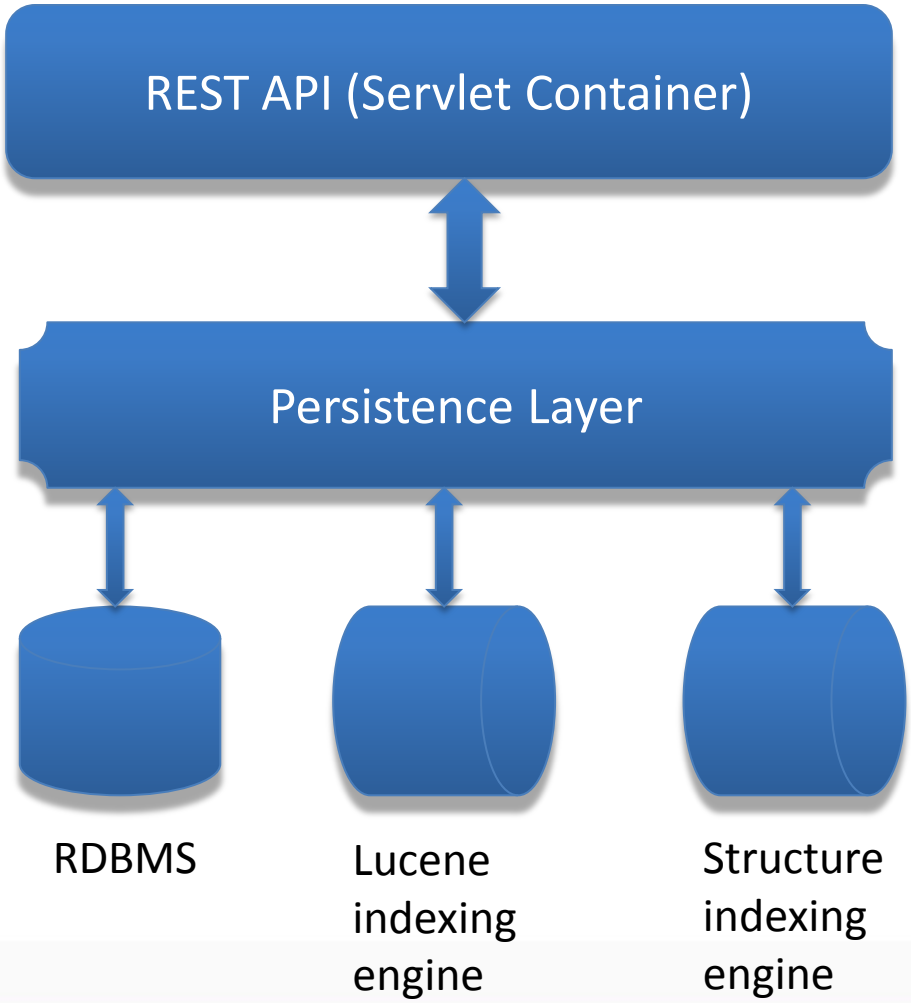
Architecture Overview

- Web-based client over RESTful interface
- RESTful API as the lowest common layer
  - » Data models and concepts based on ISO 11238
  - » Support legacy/existing database infrastructure via mapping to common data models
  - » JPA/JDO as data persistence layer for standalone instances; i.e., database vendor neutral.
- JSON as the default data exchange format; MDL's SDF as the molecular format.
  - » Specialized message format (e.g., SPL) supported on a case-by-case basis
- Deployable in any J2EE (version 5+) container
- Fine-grained access control
  - » Resource (API), domain (jurisdiction), role (e.g., analyst, administrator, etc.), field- and relationship-visibility (e.g., is the association between a particular name and structure public knowledge?)
- Backend database agnostic (e.g., Oracle, MySQL)
- Structural handling of documents
- Where possible open source is preferable

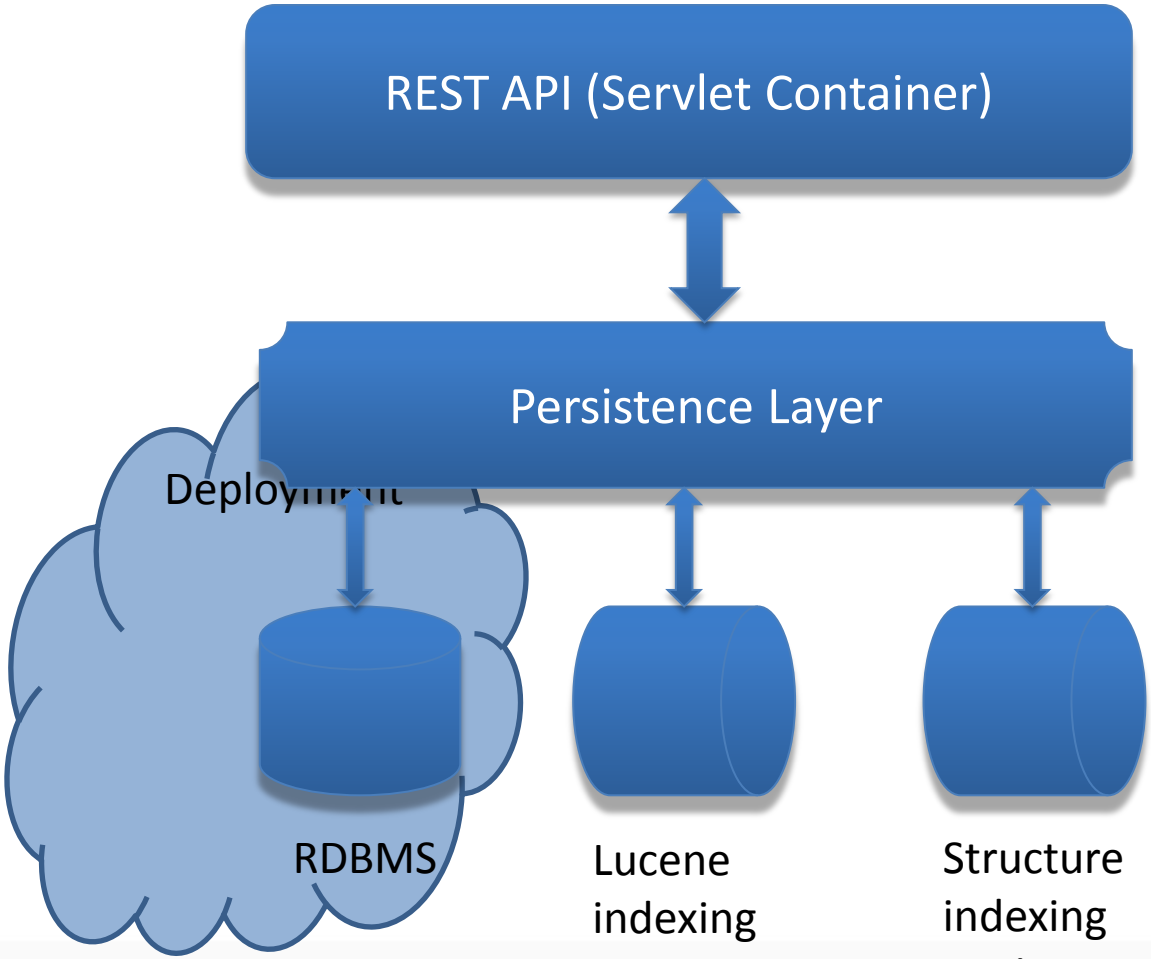
Architecture Overview

- Client-server
  - » Server is self-hosting when installed on desktop
- Backend database agnostic (e.g., Oracle, MySQL)
- Standalone server or deployable within a standard web container (e.g., Glassfish, Tomcat)
- Pluggable chemical toolkit
  - » Default toolkit is a redistributable version of JChem
- Pluggable engines for text, structure, and sequence searching

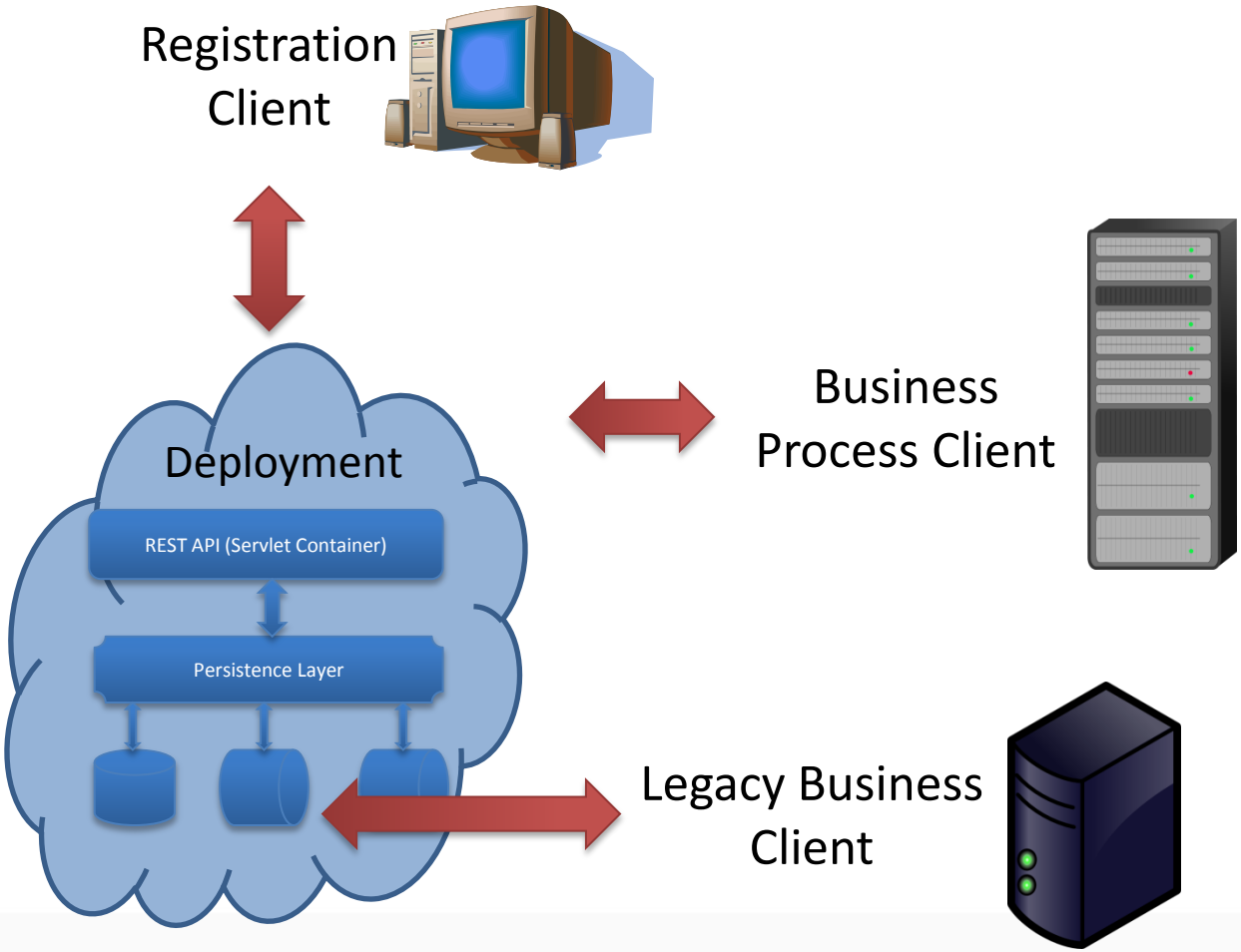
RESTful Architecture



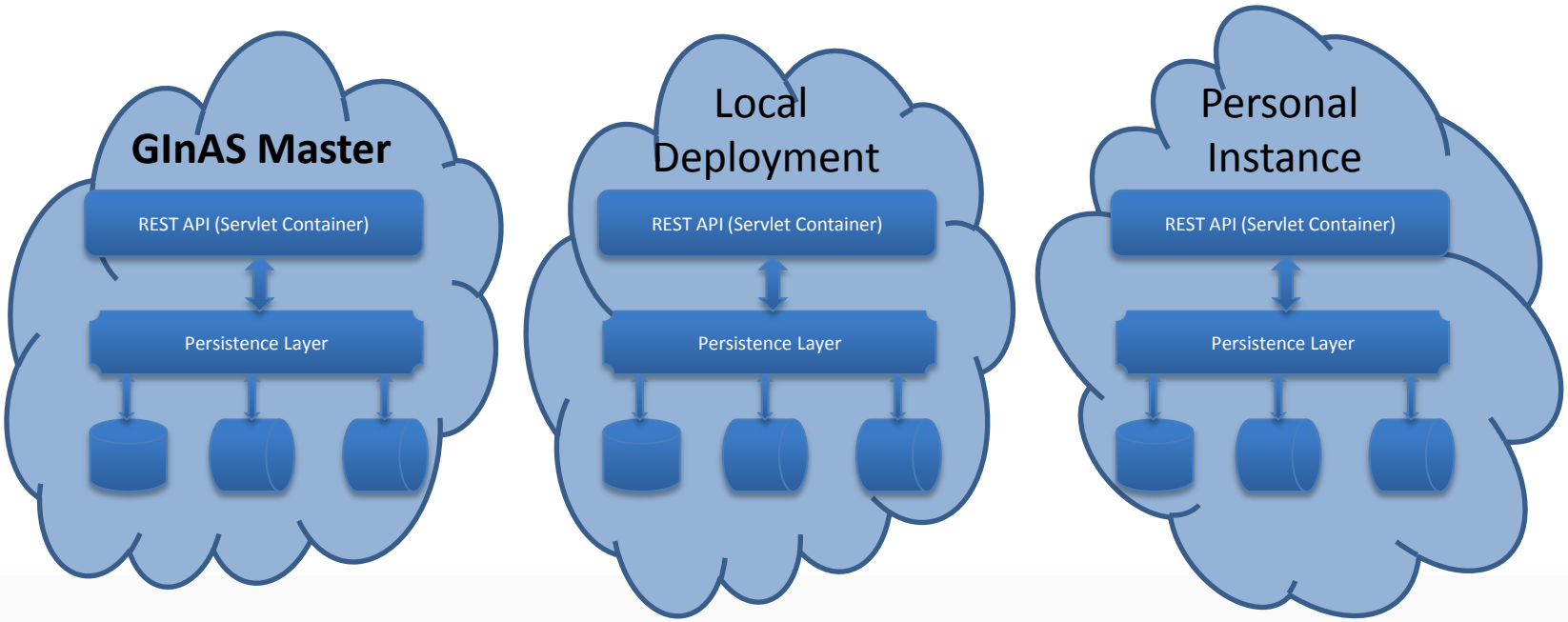
RESTful Architecture



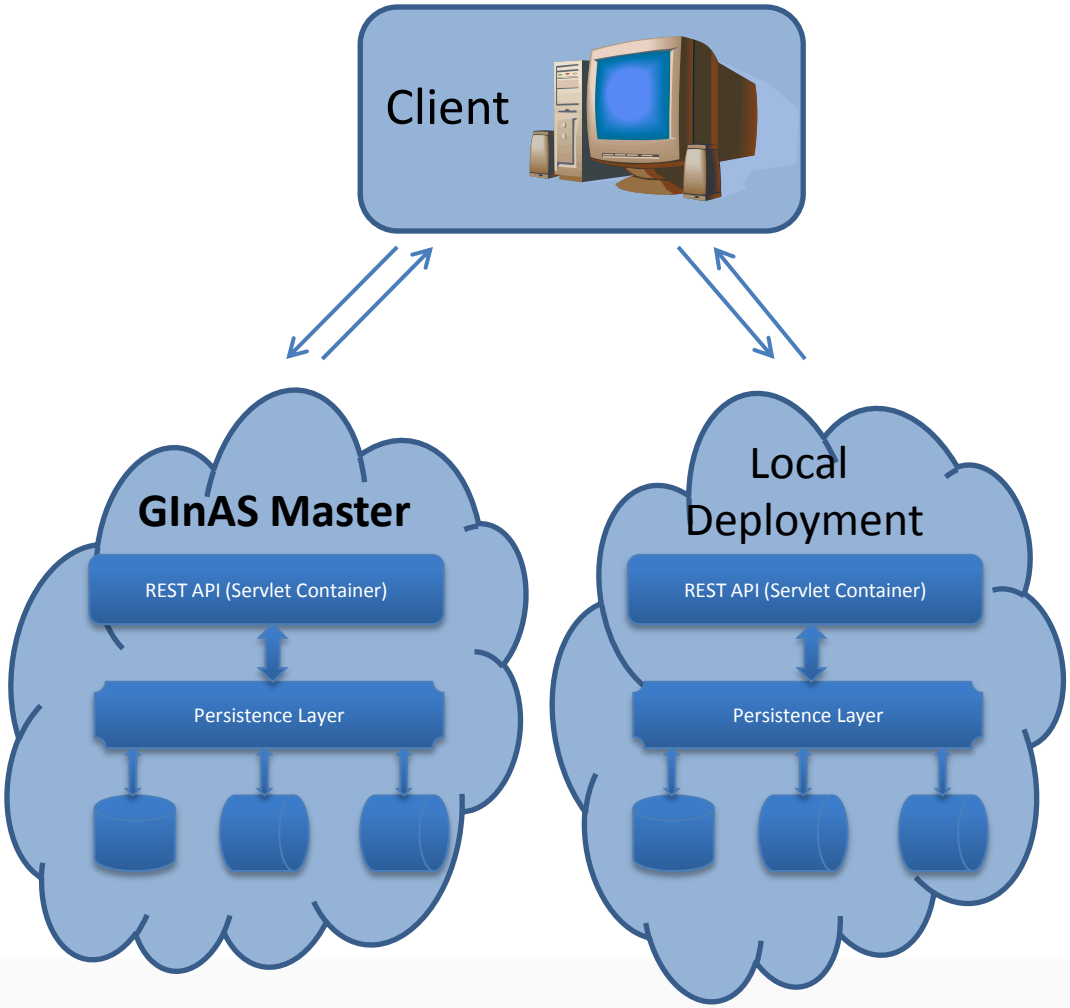
Many clients to one API instance



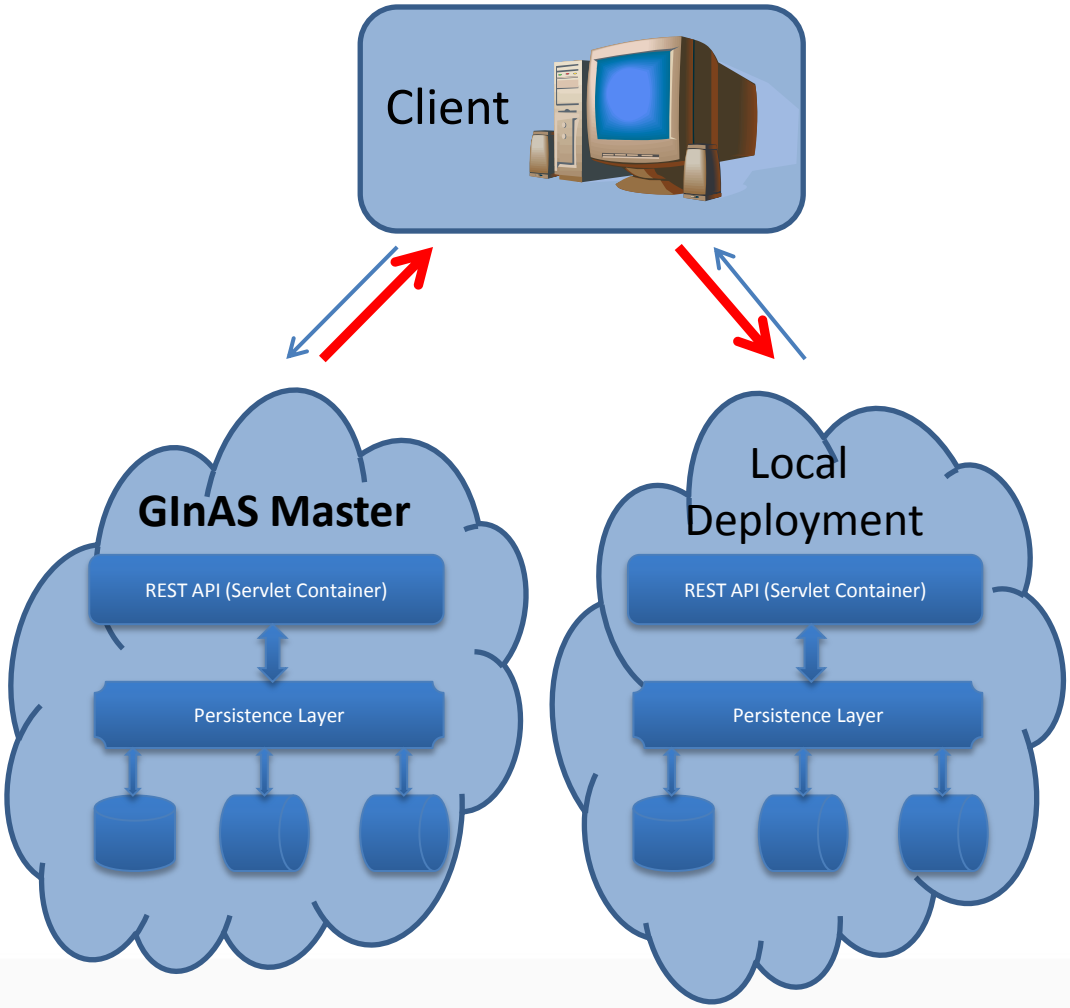
Distributable; Many API instances



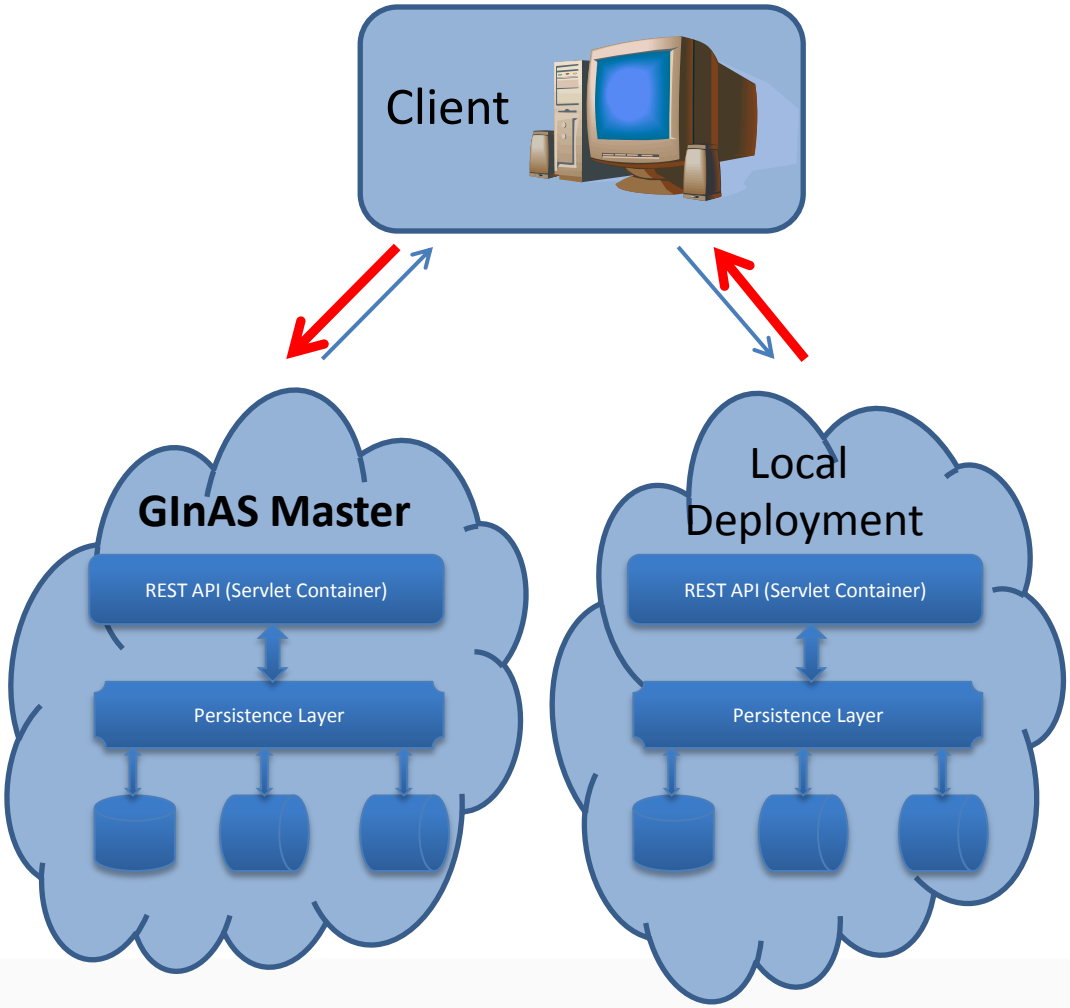
Data exchange between instances



Data exchange between instances



Data exchange between instances



Reference Substance Data

- A reference substance database is distributed with each software deployment
  - » Initial bootstrap from FDA's public SRS and NCATS data
- Instances can participate in one-way (pull) or two-way (pull/push) updates with each other
  - » Data curation
  - » Conflict resolution
- Defined update schedule

Access Control and Data Protection

- Within RDBMS integrity, security configuration
- RDBMS access exclusively through API
  - » Defined SQL protects performance, data integrity
  - » API Key level of authentication and access possible
  - » HTTPS encrypts data exchanges including credentials
- Integrate existing authentication and authorization models (e.g., single sign-on)
  - » API-level token to process transactions or
  - » Pass-through API to support LDAP with in RDBMS

Deployments will have their own security requirements and need to support a variety of options

Current System and Examples

- Health Canada Prototype

  http://ginas.hc.ircan-rican.org/

- System presentations and functional designs

  https://tripod.nih.gov/pub/ginas/
- Retrieval by name

  https://tripod.nih.gov/ginas/v6/OXYTOCIN/name
- Substructure searching

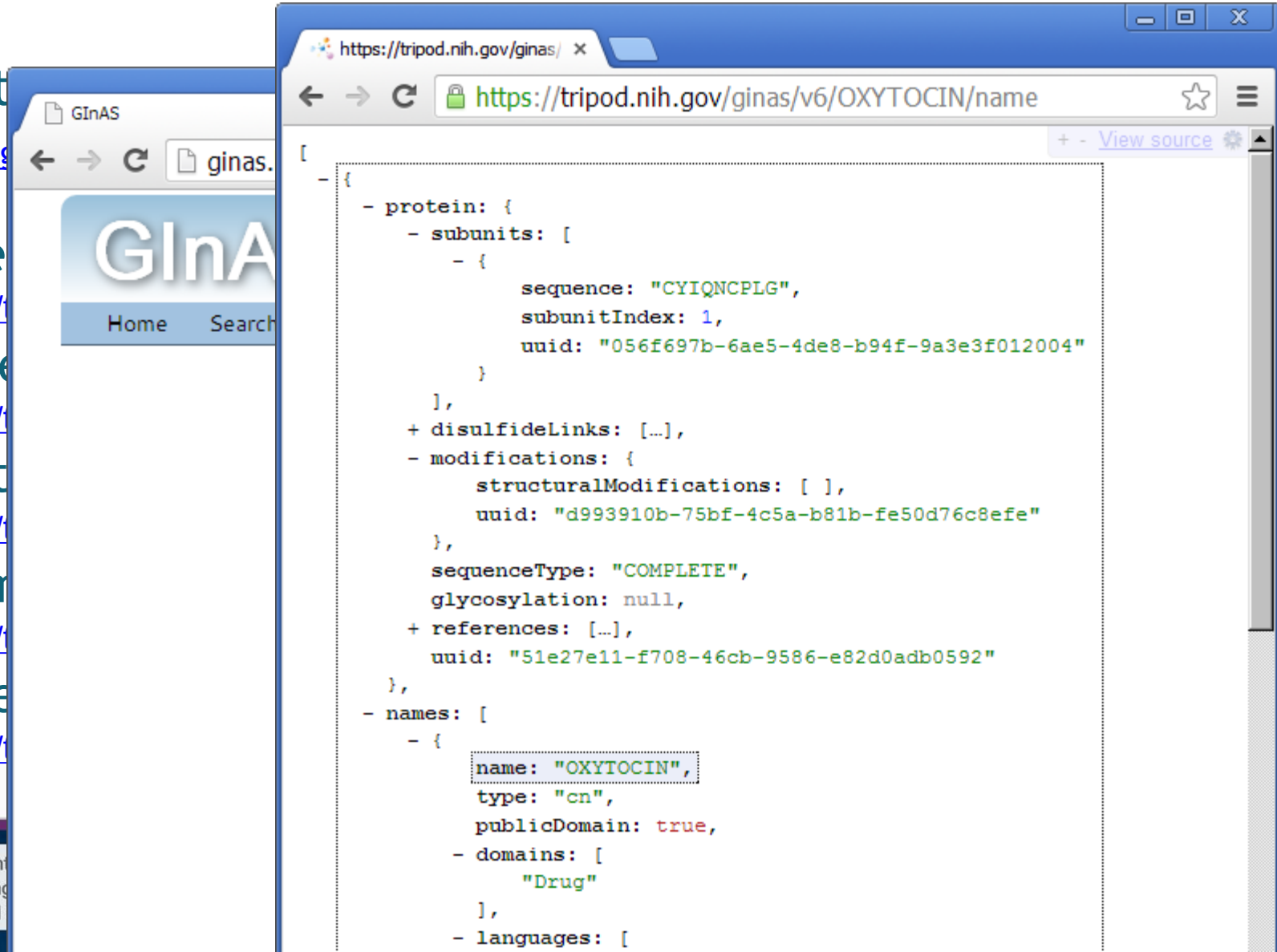
  https://tripod.nih.gov/ginas/v6/c1ncncn1/sub
- Tanimoto similarity searching

  https://tripod.nih.gov/ginas/v6/CCNC1=NC(Cl)=NC(NCC(O)=O)=N1/sim?cutoff=.75
- Name resolver

  https://tripod.nih.gov/ginas/v6/resolver/gleevec/glivec/names

Current System and Examples

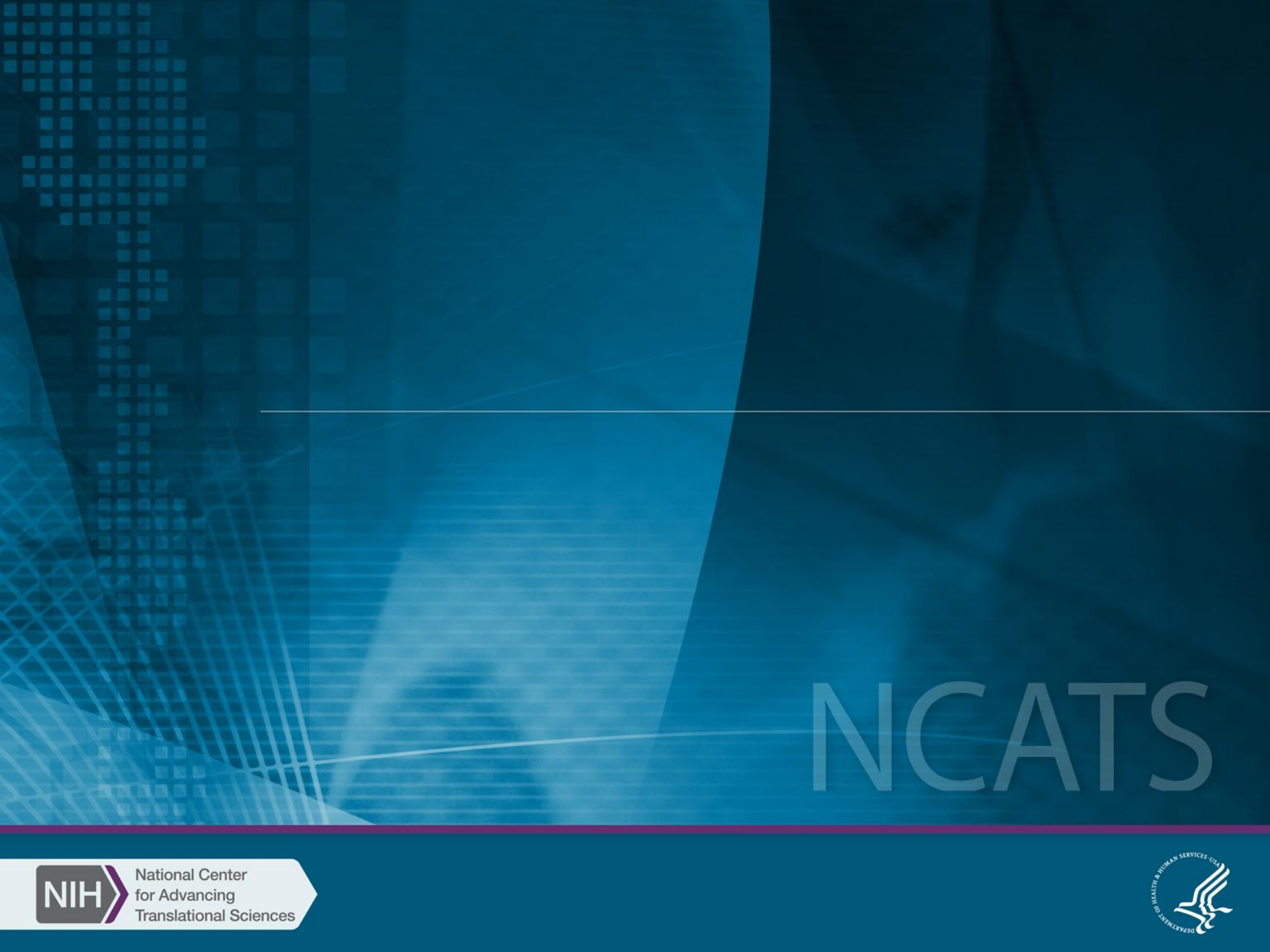
- Healt
  http://g
- Syste
  https://t
- Retrie
  https://t
- Subst
  https://t
- Tanim
  https://t
- Name
  https://t



Implementation Roadmap

Staged Development

- Stage I (alpha) http://ginas.hc.ircan-rican.org/
  - Functional design, use cases from FDA & CBG-MEB
  - Implement core services and data models for substance categories
  - HealthCanada: Host GINAS demonstration instance for substance registration, July 2013
  - FDA: Local deployment for chemicals, proteins, structural diverse migrated from FDA public SRS, October 2013
- Stage II (beta)
  - Support for other substance types, Including specified level 1 and 3, December 2013
  - Open test deployments to other organizations, April 2014
- Stage III (public release)
  - Higher levels specified substance support, June 2014
  - API supports incremental data updates; software updates bi-annually


NCATS
NIH
National Center
for Advancing
Translational Sciences